A New Perspective on Reusing Semantic Resources

Diana Maria Trandabăț *dtrandabat@info.uaic.ro*
University "Alexandru Ioan Cuza" of Iasi
Iași, Romania

Oana Gagea *oana.gagea@info.uaic.ro*
University "Alexandru Ioan Cuza" of Iasi
Iași, Romania

Abstract

Well trained linguists manage to capture semantic behavior of words in various annotated corpora. Using them as training data, semantic relations can be discovered by intelligent systems using supervised machine learning techniques. What if we have short deadlines and limited human and financial possibilities that prevent us from building such a valuable training corpus for our own language? If such a corpus already exists for any other language, we could make use of this treasure and reproduce it for the language we need.

This paper proposes an import method, which transfers semantic annotation (which could be semantic roles, named entity, sentiments, etc.) from an annotated resource to another language, using comparable texts. The case of semantic role annotation transfer from English to Romanian is discussed.

Keywords: natural language semantics, comparable corpora, semantic roles, annotated semantic resources.

1. Introduction

In order to teach computers to understand a human speech, language models need to be specified and created from human knowledge. While still far from completely decoding hidden messages in political speeches, computer scientists, electrical engineers and linguists have all joined efforts in making the language easier to be learned by machines.

A key concern in the natural language processing field is the identification of the mechanism that allows the attachment of meaning to larger chunks of text, including the study of sense and denotative references, argument structures, semantic roles, discourse analysis, and the linking of all of these to syntax.

The main research question this paper intends to answer to is up to what extent we can reuse semantic annotation. The interest begun when observing the huge amount of time and human resources involved in creating the FrameNet semantic role resources for English [3, and later for German [8], Spanish [22] and Japanese. Since semantic information is considered of major influence for a natural language processing (NLP) system, we started to consider developing such a resource for a different language, but with considerable less human and temporal resources.

The paper is structured as follow: Section 2 gives a short overview of semantic roles and the methods used in developing annotated semantic role resources for different languages, before presenting the existing approach to reproduce the FrameNet semantic resource for Romanian in section 3. Section 4 moves beyond the existing approach, by introducing a new method for semantic annotation import, which uses comparable texts instead of parallel ones. Section 5 briefly discusses the evaluation of this method, before drawing some conclusions in the last section.

2. Semantic Roles

The intuition that semantic analysis can make a positive contribution to language-based applications has motivated the development of a number of lexical-semantic resources. The potential contribution of these resources is constrained by the information they contain and the level of effort involved in their development.

Recently, there has been a growing interest in more in-depth semantic analysis for practical NLP tasks, in particular as a basis for open-domain information access. As shown in [7], NLP systems gradually stopped relying so much on word-based techniques, and started to exploit more consistently semantics. Large-scale lexical semantic resources, such as WordNet [10], have been developed and put to use for approximate semantic modeling in many applications. The lexical semantic resources developed within the FrameNet [3] and PropBank [19] projects are also valuable resources for automatically modeling the predicate-argument structure.

The semantic relations can be exemplified within the Commercial Transaction Frame, whose actors include a *buyer*, a *seller*, some *goods*, and some *money*. Among the large set of semantically related predicates, linked to this frame, we can mention *buy*, *sell*, *pay*, *spend*, *cost*, and *charge*, each of which indexes or evokes different aspects of the frame. The verb *buy* focuses on the *buyer* and the *goods*, backgrounding the *seller* and the *money*; *sell* focuses on the *seller* and the *goods*, backgrounding the *buyer* and the *money*; *pay* focuses on the *buyer*, the *money*, and the *seller*, backgrounding the *goods*; and so on. The idea is that knowing the meaning of any of these verbs requires knowing what takes place in a commercial transaction and, to some extent, knowing the meaning of all the predicates involved in the frame.

3. Transferring Semantic Annotations

Annotated language resources have become a must in natural language processing, especially for supervised learning (training and evaluation), unsupervised learning (evaluation), hand-crafted systems (evaluation), etc. Quality control is an important issue, since annotations, in order to be used as gold standard for evaluation, need to be very accurate. Inter-annotator agreement metrics have been developed (an overview is presented in [1]), but the major problems remain the temporal, financial and human resources needed in order to ensure a (near) perfect corpus. What if we have short deadlines and limited human and financial possibilities? *Could we re-use existing language resources, built with considerable effort for a specific language, and import them for a new language*? In our previous experiments [23,24] ,we proved that this approach is feasible by building a small semantic role resource for Romanian through import of English FrameNet annotation, using parallel texts.

Multilingual databases built along the FrameNet model are organizationally similar. The language specific aspects of definitions include example sentences and generalizations of syntactic realization patterns. The English examples are replaced in FrameNets for other languages by original examples from those languages that fulfill the same function.

Creating from scratch a semantic role resource implies several steps before the annotation process itself: (1) finding a corpus, (2) establishing an annotation schema and defining annotation guidelines, (3) choosing/creating annotation software, (4) training annotators. The basic daily routine of semantic role resource annotators for the English FrameNet is defined in [2] and involves:

- define a frame and its roles;
- make a list of words that evoke the frame (its lexical units - LUs);
- extract example sentences containing these LUs from a corpus;
- semi-automatically annotate the parts of the sentences which are the realizations of these roles, including marking the phrase type (PT) and grammatical function (GF).
- automatically create a report which constitutes a lexical entry for this LU, detailing all the possible ways in which the roles can be syntactically realized.

Creating a semantic role resource for a new language, as presented in [8] and [22], is usually performed similarly to the way the English semantic roles resource was created. The main difference is that the annotation schema is already established. However, it still takes considerable time to go through the annotation process on the new corpus. The starting point for the German, Japanese and Spanish FrameNet creation was the manual annotation at semantic role level of existing corpora for each language.

With the development of word alignments methods for parallel corpora, the idea of using one languages annotation to induce an annotated resource for another language has come into view. Yarowsky et al. [26] have described a system and a set of algorithms for automatically inducing stand-alone monolingual part-of-speech taggers, named-entity taggers and morphological analyzers from English to French, Chinese, Czech and Spanish. The assumption that for two sentences in parallel translation, the syntactic relationships in one language directly map to the syntactic relationships in another language (named the direct correspondence assumption) has also been studied in [11]. They provided a Direct Projection Algorithm for transferring English syntactic relations to Chinese. For the Romanian language, [4] analyses the import of verbal dependency relations on a word-aligned parallel English-Romanian corpus. After the analysis of the syntactic relations transfer, the semantic relations received the researchers' attention, especially after the development of large English resources.

The transfer of semantic information from one language to another has started to be considered for WordNet sense mapping [5], [15] and anaphora resolution [20]. At the "Romance FrameNet" Workshop held at Eurolan 2005 Summer School[30] , two papers have begun to investigate the transfer of semantic relations from English FrameNet to Spanish [13] and to Romanian [23] using word-level alignment of parallel corpora.

To save time, [24] proposed an approach which directly imports English annotation to Romanian by creating a parallel corpus through translation of the sentences from the English annotated resource. The intuition behind that approach was that most of the frames defined in the English FN are likely to be valid cross-linguistically, because semantic frames express conceptual structures, language independent, at the deep structure level.

The steps required in importing semantic roles from English to another language, as presented in [24] are: (1) select an English annotated sentence, (2) translate it to Romanian, (3) align the English and the Romanian sentence at word level, (4) transfer the annotation from English to Romanian, (5) validate and correct the import.

For the import method, the main time consuming task is the translation. After the automatic alignment (error prone) and import, a linguist performs the validation of the created corpus, focusing on cases where the alignment was not 1:1. The automatically importing program is based on the correlation of the semantic roles expressed in English with the translation equivalents in Romanian of the words that realize a specific role. The interface allows the user to perform the following tasks, related to the import of semantic roles:

- load sentences from an annotated English frame;
- translate them (either manually or using Google Translate, potentially followed by validation);
- word level alignment of the English and Romanian versions (using COWAL described in [25]);
- transfer the roles from one language to the other using a set of transfer rules;
- visualize, correct and save the annotation files.

The main drawback of this existing method lies in the time consuming and error prone translation and alignment steps. To bypass them, we propose now a new method, which builds comparable texts, instead of parallel ones, for existing English sentences and imports their roles. To our best knowledge, comparable texts have never been considered before for similar tasks.

[30] Romance FrameNet Workshop web page: http://www.icsi.berkeley.edu/ vincenzo/rfn/index.html

4. Using Comparable Texts

Comparable corpora refer to bilingual texts which are similar in form and content, yet are not translations of each other. A definition comes from [17], focusing on similar domains and sampling period:

> *"A comparable corpora can be defined as a corpus containing components that are collected using the same sampling frame and similar balance and representativeness."*

Although progress has, no doubt, been made in the construction of comparable corpora, there is very little literature in which the characteristics of comparable corpora are explained and analyzed. It would seem that they are mostly ideals rather than realities.

Since we needed a method to bypass the translation and alignment steps in the existing annotation transfer program, comparable corpora seemed to be the perfect match. However, in our case, the FrameNet annotations we wanted to import consist in a resource of separate sentences, with no connection between them. Since FrameNet is primarily a lexicographic project, only isolated sentences were considered, with the purpose to exemplify the range of combinatorial possibilities of a target predicational word. Thus, our research in identifying comparable texts was simplified at finding sentence pairs in Romanian which exhibit (at least partial) similarities in content with the English annotated sentence.

The following subsections present the architecture of our improved import program, based on comparable sentences. Our system works fully automatically, and, although we tested it on the transfer from English to Romanian, it can also be used for different language pairs.

Sentence Selection

For testing our approach, we selected 200 sentences from the English FrameNet resource, from different semantic frames, having at most 3 annotated semantic roles (see the evaluation section for discussion on the number of semantic roles). A sentence is randomly selected from this subcorpus of the English FrameNet. In this article, we will exemplify our methodology using the following annotated sentence, for the target word `hit`:

But [Cormack]$_{\text{Experiencer}}$ reeled as the ship began to roll and [hit]$_{\text{Target}}$ [his elbow]$_{\text{Body\_part}}$ [on the ladder rail]$_{\text{Injuring\_entity}}$

Identification of Predicate (Target) Word and Semantic Roles

The predicational word (the target, the word for which semantic roles are marked) and its semantic roles are extracted from the English file. In our example, the target word is the verb `hit`, with three roles: an experiencer (`Cormack`), a body part being hit (`elbow`) and an injure causing entity (`ladder rail`).

Keywords Extraction

This module extracts relevant words in the sentence by ignoring a set of stopwords. The stopwords list consist mainly of the closed classes of part of speeches, such as articles, prepositions, conjunctions, pronouns, numerals, etc. Additionally, we also included in the stopwords list auxiliary and modal verbs. The list of keywords for the analyzed sentence is: `{Cormack, hit, elbow, ladder rail}`.

These keywords are going to be used to generate the query for the search module. If they contain a named entity, the next module will be called.

Named Entities

Named Entities are very important for finding comparable sentence pairs. However, they may be spelled differently in the two languages. Finding common named entities in phrases from texts in different languages is a powerful indicator that the phrases may be translation equivalents[31].

[31] The "modulo anaphora" phenomenon [21]

In our program, named entities are simply considered words beginning with capital letters, excepting the first word in the sentence. A special module was built for tokens with capital letters which are the first tokens in phrases, following the procedure described in [12].
There are usually two mapping situations: (1) more or less phonetically equivalent named entities (e.g. cities like München – germ. / Munich – en.) and (2) named entities whose component words are translated individually (e.g. ―Black Sea –en. / Mer Noire –fr.).
In order to address named entity changes, we sequentially applied two modules. The first one considers Wikipedia page titles. Wikipedia is not a parallel corpus, as their articles are not translations from one language into another. However, Wikipedia pages are linked to their multilingual versions. Thus, using Wikipedia page titles, we can find translation of English named entities in Romanian.
The second method only applies if there is no result of the first method and it involves using transliteration similarity. Named entities are frequently not orthographically equivalent, thus transliteration from the writing system of one language to that of another is needed. Similar to existing phonetic-based approaches, such as the ones in [14] and [16], candidate matches are determined using a set of rules derived from edit distance measures (Levenstein distance).
Returning to our example, from the list of keywords, the name Cormack is selected. It is searched in the list of Wikipedia page titles. Two titles are found: *Cormack,_Newfoundland_and_Labrador* and *Cormack_(surname)*, but none are linked to Romanian versions, therefore transliteration rules are applied and the named entity list becomes: `{Cormack, Cormac}`.

Query Creation

In order to create the query, keywords are translated using Bing Translator's API [6]. For our example, the query contains the following elements:
Q = {(Cormack, Cormac), (lovi), (cot), (balustradă)}
Multiple queries are created, containing:

- The target word with all semantic roles, one variant at a time;
- Pairs of target word with different semantic roles, one by one.

Q = {(Cormack lovi cot balustradă), (Cormac lovi cot balustradă), (Cormack lovi), (Cormac lovi), (lovi cot), (lovi balustradă)}

Search and Ranking Module

All queries are searched on Google, and the first 10 results are selected for each of them. Snippets containing all keywords in one sentence are extracted, while results having the keywords in different sentences are disregarded.
In our example we had 6 queries to search over the web. The first two queries, which we named full queries (i.e. including all semantic roles) returned no results, indicating that the sequence formed by the target verb *hit* along with its three roles was not found in Romanian documents. For the third query, *Cormack lovi*, the search returned 8900 results. Out of the first 10 considered, only one contained the keyword *Cormack* and the *Target* verb *lovi* in the same sentence (example 2 below). The next query reports even more results returned, but in the first 10 considered, none contained both the *Experiencer* role *Cormac* and the *Target* verb *lovi* (en: hit) in the same sentence. Not surprisingly, the next query, *lovi cot* (en: hit elbow) returned thirty times more results, and 8 out of the first 10 contained the *Body*_part role and the *Target* in the same sentence. The sentence in example 3 below is the first one retuned, having with the higher score. The last query `lovi balustradă` also returned a significant number of results (more than 20.000), but out of the first 10, only 5 contain the *Target* word `lovi` and the *Injuring_part* role `balustradă` (en: ladder rail) in the same sentence. The first returned result was:

(1) RO: Când duba va lovi balustrada podului, vom simţi cu siguranţă.

EN[32]: We will definitely feel when the van will hit the bridge railing.
The list of snippets is further ranked using scores which consider the number of keywords found in the retrieved texts, the total number of search results provided by Google for the specific query and the rank in the 10 search results. From the ranked list of snippets, only two[33] are retained and annotated.

Annotating Semantic Roles

The top two sentences retained from the previous step are:
(2) RO: John **[Cormack]**$_{\text{Experiencer}}$ putea fi **[lovit]**$_{\text{Target}}$ atât de adânc, încât să răspundă pozitiv apelurilor de a pleca de la Casa Albă.
EN: John [Cormack]$_{\text{Experiencer}}$ could be [hit]$_{\text{Target}}$ so deeply that he would respond positively to the requests to leave White House.
(3) RO: M-am **[lovit]**$_{\text{Target}}$ acum 3 saptamani la **[cot]**$_{\text{Body\_part}}$.
EN: I [hit]$_{\text{Target}}$ my [elbow]$_{\text{Body\_part}}$ 3 weeks ago.
Even if no sentence was retrieved containing all the searched semantic roles, the two sentences above can be automatically annotated with the corresponding semantic roles. This is easily performed by preserving, for each keyword in the query, a pointer to its initial semantic role from the English annotation. Thus, the output of our proposed annotation import program is, for each input sentence from the FrameNet resource, a Romanian sentence annotated with at least one of its semantic roles.

5. Evaluation

Validation was manually performed. At this stage, we are still fine-tuning our improved import program, so we only validated a small number of imports (200 sentences). During the validation process, several problems occurred, which will be detailed in this section.

For the evaluation of our import program based on comparable sentences, we only imported roles for sentences having at most 3 annotated semantic roles in the English version. This limitation[34] was introduced in the testing version of our system due to the observation that only a very small fraction of the snippets retrieved for the full query included all semantic roles, and most of them were incomplete sentences.

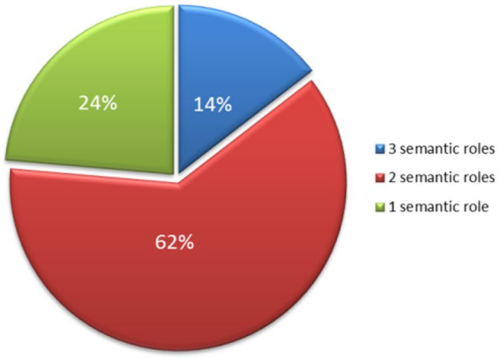

Fig. 1. Number of semantic roles per predicate

Figure 1 gives an overview of the proportion of input sentences with different number of annotated roles in the English subcorpus of FrameNet that we considered as input for our import program.

[32] Manual translations of Romanian examples are given for the purposes of this paper and are marked with "EN:" in sections 4 and 5.

[33] This limit is set in order to avoid, as much as possible, noise introduced by the search engine.

[34] This limitation can be removed by choosing a different parameter for our import program at runtime.

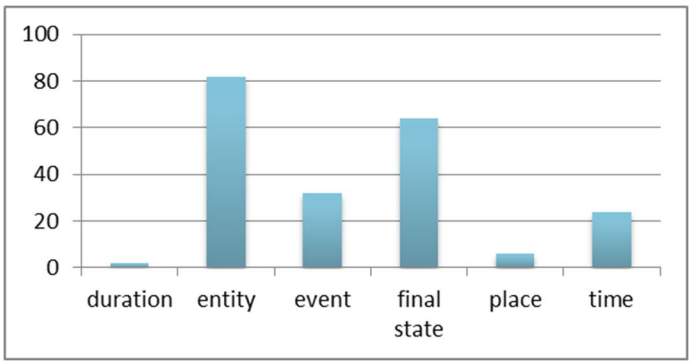

Fig 2. Distribution of the different types of imported semantic roles

The 200 sentences we considered as input corpus contained annotated semantic roles extracted from the English FrameNet for the *Event* and the *Becoming* semantic frames. Six semantic roles are marked in our input corpus for the two semantic frames: *Entity*, *Event*, *Final State*, *Duration*, *Place* and *Time*.

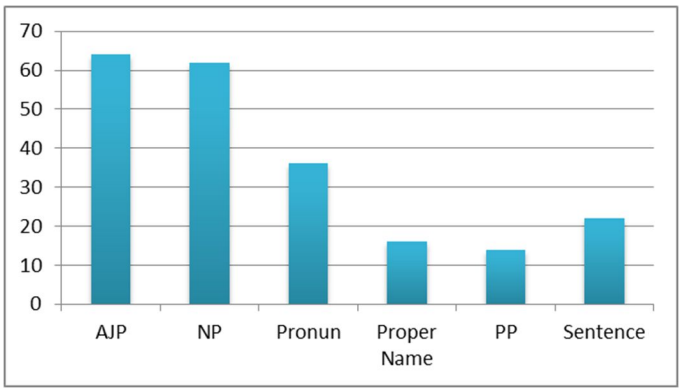

Fig. 3. Distribution of phrase types

Figure 2 presents a distribution of the types of semantic roles found in the English subcorpus of FrameNet, while figure 3 presents the different phrase types that the English annotated semantic roles have. The considered phrase types are as follow: AJP – adjectival phrase, NP – noun phrase, Pronoun, Proper Name, PP – prepositional phrase, Sentence.
The evaluation of the success of our program was computed using the distribution of semantic role types and their phrase type. Additionally, the average length of the phrase type seemed to influence the import results (see table 1).

Table 1. Success rate for different phrase types

Phrase type for English semantic roles	Average length (no. of words)	Successful imports (%)
AJP	4.57	75.00
NP	6.27	48.38
Pronoun	4.00	94.44
Proper Name	2.50	87.50
PP	8.00	14.28
Sentence	12.57	9.09

For instance, *Final_state* semantic roles usually contain long nominal or adverbial phrases, as well as very long relative sentences. Thus, the probability of finding a relevant snippet to extract a comparable sentence from it decreases as the size of the role in English increases.
As an example, the program finds no results for the *Target- Entity-Final_State* pair:

[The lengths to which he is prepared to go]$_{Entity}$ will [become]$_{Target}$ [clear]$_{Final\_State}$ [when he and Esau meet]$_{Time}$.

but found more than 1500 results for the exact *Target- Entity-Final_State* pair in the next example:

[He]$_{Entity}$ [became]$_{Target}$ [nervous]$_{Final\_State}$.

Table 2 presents the success rate for the correct import of different semantic roles. One can notice that all roles tagged with the *Place* tag were correctly transferred, mainly due to the fact that they were expressed by noun phrases consisting of 2-3 words. On the other hand, all roles marked with the *Duration* tag were not found by the web search module.

Table 2. Percentage of successfully imported roles

Semantic role in English	Successful imports (%)
Duration	0.00
Entity	61.90
Event	62.50
Final state	56.25
Place	100.00
Time	27.27

The motivation is due to the length of the constituent (when he and Esau meet) or the fact that it represents an interval, hard to be found as such in a web search (between October 29 and November 3).
When evaluating the performance of the improved import program, we identified 3 main error cases, discussed below.

Ambiguity of Semantic Roles

The surface realization of the semantic roles may have different senses. For instance, consider the example:
Cormack hit his [head]$_{body\_part}$.
When translating the body part role, the Romanian equivalent cap has several word senses, some of them similar to English, others not. Thus, in the retrieved snippets, we also found the sense cape as in the example below.
RO: Un uragan a lovit Insulele **Capului** Verde.
EN: A hurricane hit the island nation of **Cape** Verde.

Systemic Order

The position of the verbal direct and indirect arguments within the *systemic order* [9] is important for annotation transfer. Consider, for instance, that we would have Paris instead of Cormack as named entity in the input sentence:
But [Paris]$_{Experiencer}$ reeled as the ship began to roll and [hit]$_{Target}$ [his elbow]$_{Body\_part}$ [on the ladder rail]$_{Injuring\_entity}$
The full query would have been Q= {Paris hit elbow ladder rail}, which returns among other the result:
In [Paris]$_{Experiencer}$ m-am [lovit]$_{Target}$ crunt [la cot]$_{Body\_part}$.

I hit my elbow brutally in Paris.

It is obvious in this situation that Paris should have been marked as *Location*, and not *Experiencer*.

Multiple Predicates

The major problem we found is the fact that semantic roles may be expressed, in comparable texts, for a different predicate. For instance, in the example below, the word `balustrada` (en: ladder rail) belongs to the verb `fall`, while the *Injuring_entity* semantic role for the verb `hit` is unexpressed.
RO: O adolescentă a căzut peste o balustradă şi s-a lovit la cap
EN: A teenager felt over a ladder rail and hit her head.
In order to limit this problem, we consider modifying our scores to include also a restriction on the number of predicates.

6. Conclusions

The major contribution of this paper is the language independent method for importing semantic role annotation from one language to another, based on automatically obtained comparable texts. Thus, we propose reusing the annotation from a semantic corpus to automatically build new annotated resources, and exemplify by developing a FrameNet-like resource for Romanian.

Our results suggest that semantic role information can be automatically imported from one language to another, with various degrees of success for different phrase types and imported roles. This pilot study need to be extended to a larger scale, considering also other types of semantic roles and phrase types.

The semantic concept level is mostly cross-linguistically constant, the surface realization of these concepts and their relations (the syntactic constraints) being changed from one language to another. However, our method is definitely better than the one using parallel texts, since the English-centric approach is avoided. In the previous approach, the realization of a language specific FrameNet could be endangered by tempting the exact copying of the English semantic frames. In the present approach, inter-lingual differences may be actually manifested, since texts are extracted from genuine Romanian web pages. After all, the transfer starts from lexical predicates, and lexicals are language dependent.

As further work, we intend to investigate sentences with technical terms [21], which may also suggest similarities of sentences. Another promising method for identifying similar sentence pairs within comparable corpora, proposed in [18], will also be investigated.

References

1. Artstein, Ron and Massimo Poesio. Bias decreases in proportion to the number of annotators. In: Proceedings of FG-MoL2005, pp 141–150, Edinburgh (2005)
2. Baker Collin F. and Hiroaki Sato. The FrameNet data and software. Poster and Demonstration at ACL2003 (2003)
3. Baker Collin F., Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet project. In: Proceedings of the COLING-ACL1998 (1998)
4. Barbu Mititelu Verginica and Radu Ion. Automatic Import of Verbal Syntactic Relations Using Parallel Corpora. In: Proc. of the RANLP, pp 329–333 (2005).
5. Bentivogli Luisa and Emanuele Pianta. Opportunistic Semantic Tagging. In: Proc. of LREC '02, pages 1401-1406, Las Palmas, Canary Islands-Spain (2002)
6. Bing Translator API (2016), https://www.microsoft.com/en-us/translator/translatorapi.aspx Accessed April 16, 2016
7. Cambria E. and B. White. Jumping NLP curves: A review of natural language processing research. IEEE Computational Intelligence 9(2), pp. 48-57 (2014)
8. Erk Katrin, Andrea Kowalski, Sebastian Pado, and Manfred Pinkal. Towards a resource for lexical semantics: A large German corpus with extensive semantic annotation. Proc. of ACL03, Sapporo, (2003)

9. Eva Hajičová, H. Skoumalova, and P. Sgall. An Automatic Procedure for Topic-Focus Identification. Computational Linguistics, 21 (1): 81-94 (1995)
10. Fellbaum, Christiane. WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670 (2005)
11. Hwa Rebecca, Philip Resnik, Amy Weinberg, and Okan Kolak. Evaluating translational correspondence using annotation projection. In: Proceedings of ACL '02, pp 392–399 (2002)
12. Iftene, A., Trandabăţ, D., Toader M., Corîci, M. Named Entity Recognition for Romanian. In: Studia Universitatis, Babes Bolyai University Publishing House, Volume LVI, Number 2, pp. 19-24 (2011)
13. Johansson R. and P. Nugues. Using Parallel Corpora for Cross-language Projection of FrameNet Annotation. In: ROMANCE FrameNet Workshop at EUROLAN 2005 Summer School, Cluj-Napoca, Romania (2005)
14. Kondrak, G. A New Algorithm for the Alignment of Phonetic Sequences. In: Proc. NAACL'00, pp. 288-295 (2000)
15. Lupu Monica, Diana Trandabat, and Maria Husarciuc. A Romanian SemCor aligned to the English and Italian MultiSemCor. In: 1st ROMANCE FrameNet Workshop at EUROLAN 2005, pp 20-27 (2005)
16. Mani, I., Yeh, A., Condon, S. Learning to Match Names Across Languages. In Proc. of COLING 2008 Workshop on Multi-source Multilingual Information Extraction and Summarization MMIES'08, pp 2-9 (2008)
17. McEnery, A. Corpus linguistics. In R. Mitkov (ed.) Oxford handbook of computational linguistics (pp.448-63). Oxford University Press (2003)
18. Munteanu, D. and Marcu, D. Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. Computational Linguistics, 31(4) (2005)
19. Martha Palmer, Dan Gildea, Paul Kingsbury, The Proposition Bank: A Corpus Annotated with Semantic Roles Computational Linguistics Journal, 31:1, (2005)
20. Postolache O. Cristea D., Orasan C.. Transferring Coreference Chains through WordAlignment. In Proc. LREC2006, (2006)
21. Skadiņa, I., Vasiļjevs, A., Skadiņš, R., Gaizauskas, R., Tufiş, D., & Gornostay, T. Analysis and evaluation of comparable corpora for under resourced areas of machine translation. In 5th Wksh on Building and Using Comparable Corpora (2012)
22. Subirats-Ruggeberg Carlos and Miriam R. L. Petruck. Surprise: Spanish FrameNet! In Wksh on Frame Semantics, Prague, Czech Republic, (2003)
23. Trandabat D., Husarciuc M., Lupu. Towards an automatic import of English FrameNet frames into the Romanian language. In ROMANCE FrameNet Wksh at EUROLAN 2005 Summer School, pp. 28-36, Cluj-Napoca, Romania (2005)
24. Trandabăţ Diana. Towards automatic cross-lingual transfer of semantic annotation, in RJCRI 2011, ISBN 978-2-35768-024-1, pp. 403-408 (2011)
25. Tufis D, Radu Ion, Alexandru Ceausu, and Dan Stefanescu. Combined aligners. In Proc. ACL2005 Workshop on "Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond" (2005)
26. Yarowsky D. Ngai G., Wicentowski R. Inducing Multilingual Text Analsis Tools via Robust Projection across Aligned Corpora. In Proc. HLC2001 (2001)